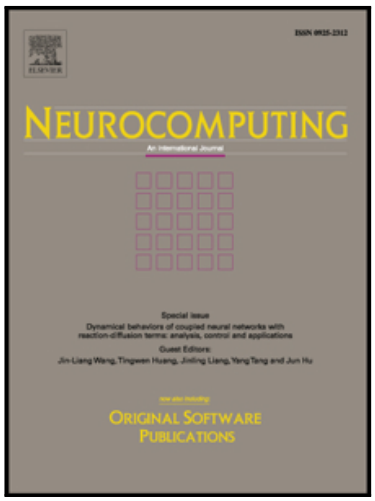

Ensemble Application of Convolutional Neural Networks and Multiple Kernel Learning for Multimodal Sentiment Analysis

Soujanya Poria, Haiyun Peng, Amir Hussain, Newton Howard, Erik Cambria

PII: S0925-2312(17)30202-3
DOI: 10.1016/j.neucom.2016.09.117
Reference: NEUCOM 18001

To appear in: *Neurocomputing*

Received date: 29 September 2015
Revised date: 4 August 2016
Accepted date: 22 September 2016

Please cite this article as: Soujanya Poria, Haiyun Peng, Amir Hussain, Newton Howard, Erik Cambria, Ensemble Application of Convolutional Neural Networks and Multiple Kernel Learning for Multimodal Sentiment Analysis, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2016.09.117

Ensemble Application of Convolutional Neural Networks and Multiple Kernel Learning for Multimodal Sentiment Analysis

Soujanya Poria[a], Haiyun Peng[b], Amir Hussain[a], Newton Howard[c], Erik Cambria[b,d]

[a]*Department of Computing Science and Mathematics, University of Stirling, UK*
[b]*School of Computer Science and Engineering, Nanyang Technological University, Singapore*
[c]*Computational Neuroscience and Functional Neurosurgery, University of Oxford, UK*
[d]*Corresponding author*

Abstract

The advent of the Social Web has enabled anyone with an Internet connection to easily create and share their ideas, opinions and content with millions of other people around the world. In pace with a global deluge of videos from billions of computers, smartphones, tablets, university projectors and security cameras, the amount of multimodal content on the Web has been growing exponentially, and with that comes the need for decoding such information into useful knowledge. In this paper, a multimodal affective data analysis framework is proposed to extract user opinion and emotions from video content. In particular, multiple kernel learning is used to combine visual, audio and textual modalities. The proposed framework outperforms the state-of-the-art model in multimodal sentiment analysis research with a margin of 10-13% and 3-5% accuracy on polarity detection and emotion recognition, respectively. The paper also proposes an extensive study on decision-level fusion.

1. Introduction

Subjectivity detection and sentiment analysis consist of the automatic identification of the human mind's private states, e.g., opinions, emotions, moods, behaviors and beliefs [1]. In particular, the former focuses on classifying sentiment data as either objective (neutral) or subjective (opinionated), while the latter aims to infer a positive or negative polarity. Hence, in most cases, both tasks are considered binary classification problems.

To date, most of the work on sentiment analysis has been carried out on text data. With a videocamera in every pocket and the rise of social media, people are now making use of videos (e.g., YouTube, Vimeo, VideoLectures), images (e.g., Flickr, Picasa, Facebook) and audio files (e.g., podcasts) to air their opinions on social media platforms. Thus, it has become critical to find new methods for the mining of opinions and sentiments from these diverse modalities. Plenty of research has been carried out in the field of audio-visual emotion recognition. Some work has also been conducted on fusing audio, visual and textual modalities to detect emotion from videos. However, a unique common framework is still missing for both tasks. There are also very few studies combining textual clues with audio and visual features. This leads to the need for more extensive research on the use of these three channels together. This paper aims to solve the two key research questions given below -

- Is a common framework useful for both multimodal emotion and sentiment analysis?
- Can audio, visual and textual features jointly enhance the performance of unimodal and bimodal emotion and sentiment analysis classifiers?

Studies conducted in the past lacked extensive research [2, 3, 4] and very few of them clearly described the extraction of features and fusion of the information extracted from different modalities. In this paper, we discuss the

Email addresses: sp47@cs.stir.ac.uk (Soujanya Poria), peng0065@ntu.edu.sg (Haiyun Peng), ahu@cs.stir.ac.uk (Amir Hussain), newton.howard@nds.ox.ac.uk (Newton Howard), cambria@ntu.edu.sg (Erik Cambria)